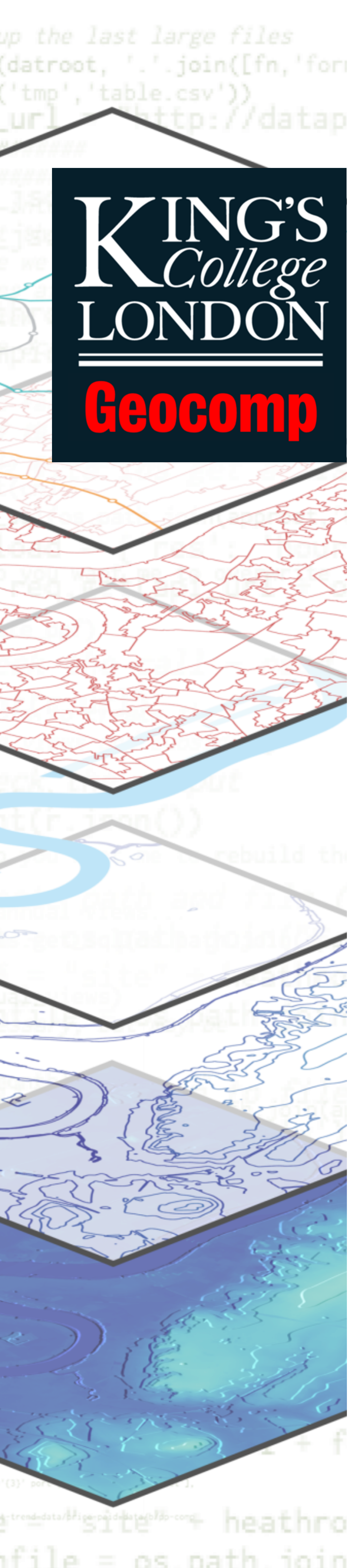
DOI:
10.1111/gec3.12403

Document Version
Early version, also known as pre-print



Geography and Computers: Past, present, and future

Dani Arribas-Bel[1] and Jon Reades[2]

1. Department of Geography and Planning, University of Liverpool, Jane Herdman Building, Liverpool L69 3GP, United Kingdom

2. Department of Geography, King's College London, Bush House (North East Wing), Strand Campus, 30 Aldwych, London WC2B 4BG, United Kingdom

This is the pre-peer reviewed version of the following article: Arribas-Bel, D. and Reades, J. (2018), 'Geography and Computers: Past, present, and future', *Geography Compass*, which has been published in final form at [https://doi.org/10.1111/gec3.12403].

Abstract

The rise of 'big data' in academia and industry has triggered something of an identity crisis for Geography: geographers are both wary of a return of 'positivist' approaches, but also excited by the potential to revolutionise our understanding of society and space. Although a detailed history of the long, and sometimes fractious, relationship between (Human) Geography and computers is beyond the scope if this article, we argue that some historical perspective allows us to better understand the current state of affairs and scope for future developments. Following a short review of the history of computation *in* Geography, we then document recent developments *outside* Geography that are reshaping our understanding of the world through data, and conclude with a reflection on how a Geographic Data Science might provide a foundation for further development.

Introduction

> *It is not the use of computers that distinguishes the forthcoming revolution but the development of a new computationally intensive and totally computer-dependent paradigm in geography.*
>
> *Stan Openshaw (1994)*

The rise of 'big data' in academia and industry has triggered something of an identity crisis for Geography: on the one hand, geographers are wary of a return of the 'positivist' approaches of the quantitative revolution of the 1960s and 70s, while on the other hand they are often excited by the scope of this new data to revolutionise our understanding of society and space. However, a wider view of our disciplinary history can help to contextualise both viewpoints as part of a much longer debate concerning the role of computation as a tool for geographical research. This article has three distinct goals: first, to quickly review the history of computation *in* Geography so as to provide a context to contemporary debates; second, to document recent developments *outside* Geography that are reshaping our understanding of the world through data; and third, to *reflect* on how a Geographic Data Science might provide a foundation for further development.

A (very) brief history

> *The key question [...] is whether [this] is to be understood as a new perspective or paradigm in geography and related disciplines, or as a grab-bag of useful computer-based tools… The question whether or not we are witnessing the rise of a distinct intellectual approach to the study of geographical space through computation…*

Adapted from Helen Couclelis (1998, p.18)

Although a detailed history of the long, and sometimes fractious, relationship between (Human) Geography and computers is both beyond the scope of this article and has been done elsewhere before (*e.g.* Armstrong, 2000; Torrens, 2005; Haining, 2013), it is useful to provide some historical perspective so as to better understand the current state of affairs and scope for future developments. Significantly, although Geography and the affiliated domain of Planning were amongst the earliest adopters of computers in the 1950s and 60s, they were also (in Britain at least) amongst the disciplines that turned most strongly *against* their use as a tool for thinking about and analysing space a few decades later. This development took place as part of what is known as the 'cultural turn' of the 1970s and 80s, and in Geography it was characterised by a range of cutting critiques (*e.g.* Ley and Samuels 1978 and see also Barnes 2004), including perhaps most notably the 'Damascene conversion' of David Harvey (Harvey1972).

There are reasons to believe the pendulum has recently begun to swing back, and a new appreciation of quantitative approaches in geography is taking shape. An important explanatory factor behind this shift resides in a series of technological advances over the last decade that is reshaping how we employ and understand computation and computers in almost every aspect of human life. The declining size and cost of chips, storage and geospatial technology has given rise to new sources of data about the world and the possibility of using them to provide new answers to old and entirely new geographical questions. Consequently, to understand the (re)emergence of computing in Geography is to understand the effects that the embedding of computers in every facet of daily life is having on social science research as a whole.

The First Wave: a computer in every institution

As early as 1963, Burton was arguing that the *first* quantitative revolution was a theoretical one and *not* a methodological one (Burton, 1963). The vanguard of this revolution saw statistics as a tool with which to uncover spatial structure, arguing that without 'observation and description of regularity' there would be nothing against which to measure – and judge – the unique and the exceptional. In his historical work on the discipline Barnes (Barnes, 2013; Barnes, 2014) echoes this view, suggesting that the work begun by, for instance, Brian Berry at the University of Washington encouraged a major shift towards the use of statistics as a tool for theory-validation. And, despite the subsequent 'cultural turn', quantitative methods *did* spread from the select few journals and departments of the early years documented in Barnes (Barnes, 2004), and carved out a permanent, if seemingly more marginal, place of their own in the discipline. For example, the flagship journal *Progress in Human Geography* (PiHG), begun in 1977, included a series of reports on quantitative methods right from the start, including advances in time series analysis (Cliff, 1977), spatial diffusion (Cliff, 1979), and modelling (Cliff, 1980).

It is neither easy, nor particularly useful, to separate this theoretical shift from the technological changes that made it possible: although most of the computation done at the time could still, in principle, be carried out by human 'calculators', the punch card and magnetic tape made it possible to do matrix manipulation and other demanding tasks at seemingly breakneck speed (*e.g.* Goddard, 1970) and consequently had a significant effect on the adoption of quantitative methods in Geography. Of course, computers at this time were large. Very large. They were expensive and hard to operate too. These constraints meant that geographers were forced to share the few machines available on campus, and this made clock cycles and computation time a precious luxury not to be wasted. A good example of the consequences these limitations imposed can be found in the numerous shortcuts, simplifications, and assumptions that fill appendices in statistical papers from those years with the goal of obtaining "computational feasibility" (*e.g.* Cliff & Ord, 1981). The computer in those days was an exciting new tool for statistical analysis at scale, but it would not be unfair to characterise it as largely substituting for the time and energy of users in the midst of a more theoretical project.

The Second Wave: a computer in every office

Without wishing to suggest that the *next* wave of innovation in computing *determined* the accompanying transformation of – and, ultimately, divisions within – quantitative geography, the growing availability of desktop computers in the 1980s inevitably had a profound effect on how we 'do geography with computers' (Harris *et al.* 2017). The dedicated desktop computer enabled the design and use of much more computationally demanding methods, perhaps most notably the development of 'local statistics' in the 90s (Haining, 2014). Poon (2003) has argued that spatial statistics can be seen as an empirical response to the critique of the cultural geographers because it explicitly incorporates variation over space. The desktop system is also, of course, intimately bound up in the rise of Geographic Information Systems (Goodchild & Haining, 2004) and, consequently, of Geographic Information Science (Goodchild, 1991).

The key point is that, with a computer on every geographer's desk, the discipline quickly began to imagine new ways to use them. The cumulative impact that the explosion of computing power was having on the discipline was summarised in the three-part series for PiHG that Stewart Fotheringham wrote exploring the local (Fotheringham, 1997), the computational (Fotheringham, 1998), and the visual (Fotheringham, 1999). Well before that, however, the *Progress* reports had already highlighted developments in discrete choice modelling (Wrigley, 1982), longitudinal data analysis (Wrigley, 1986), and input-output analysis (Thomas, 1990). This is also the period where Agent-Based Models and Cellular Automata (O'Sullivan, 2008, Torrens, 2010) emerge as a distinct path in geographical model development for exploring, principally, complexity.

Of course, in many respects the 1990s are usually seen as the decades of GIS, with a new 'reports' series in PiHG focussed solely on this approach starting in 1995. Chronicling the fast evolution of the nascent field, they explored issues in the representation, storage and analysis of spatial data. The first two reports by David Unwin covered uncertainty (Unwin, 1995) and the relation between GIS and spatial statistics (Unwin, 1996). The topics that featured most prominently during the latter part of this period were connected to challenges in data infrastructures (Longley, 2003), representation (Longley, 2004), time (O'Sullivan, 2005), and geovisualization (Elwood, 2009, 2010).

However, in Couclelis (1998, p.19) the term 'geocomputation' is used in a way that seeks to distinguish it from the dominant GIS/GISci discourse (Openshaw, 1994; Openshaw & Abrahart, 2000; Fotheringham, 1998; Haining (2014)), and she defines GIS as "a technological advance that would allow applied geographers and others to do faster, more comfortably, and better what they had always done." In other words, GIS can be seen as a ramping up of the process – doing much more quickly what was once done painfully by hand – begun in the first wave but is not, in and of itself, a form of computational *thinking*. This distinction, noted by the geography community of the time and manifested in, for instance, the neglect of topics such as Artificial Intelligence (AI) is hardly coincidental (see, *e.g.*, Goodchild, 2010). We feel that geocomputation should be seen as part of a separate tradition much more concerned with what computers make possible, not what they make easier, and that this is part of an ongoing disciplinary dialogue.

The Third Wave: a computer in every *thing*

By now it should be clear that the embedding of computational power in everyday objects, not just dedicated computers, heralds another major shift for computational geographers. Part of the significance of this third shift lies in the vast amount of affordable computational power available to store, process, and analyse an ever growing amount of data, but much of our attention has been focussed on the outputs – sometimes termed the 'data exhaust' (Harford, 2014) – of this embedding process with less attention given to the context of this change. Put simply, computers are no longer just machines with which we 'ingest' and process observations, they have become 'autonomous data generators' in their own right whose interactions and communications spawn data at volumes that dwarf our own (human) intentional generation and consumption of information. The deluge of 'big data' is therefore inseparable from a confluence of two critical trends: the declining size and cost of hardware, and the declining cost of software.

It is now possible to make cellular network-enabled devices so small and so cheap that they are, literally, disposable in the name of research. As an illustration, Phithakkitnukoon, 2013 discuss a project where dozens of customised mobile phone chips were attached to trash to track their movement through the global waste collection pipeline, with pieces of rubbish travelling across America and even internationally! Sensors are now *everywhere*: in our phones and homes, in our bridges and tunnels, orbiting the Earth in the form of nano-satellites, and (implicitly) in the digital traces that we leave in the networks with which we interact. Thanks to the rise of affordable, low-power hardware platforms such as Arduino (www.arduino.cc), as well as cheap 'self-replicating' 3D printing systems (*e.g.* reprap.org) that enable customised parts to be quickly manufactured on-site, a wealth of innovative applications in geographical data collection, particularly in the developing world, are now emerging.

The physical devices that sustain this revolution are not only cheaper because of reduced costs in materials, sensors and chips, but also because they are more accessible: the second critical trend is the expansion of 'cheap' – as in free – software. Although the desktop era was largely dominated by proprietary software running on proprietary platforms, 'free software' had been around since the early mainframe days and with the rise of Linux the use of open source code has increased exponentially and generated an entire ecosystem of freely downloadable and (re)usable software. As the first quantitative *Progress* report in ten years notes (Brunsdon, 2016), the shift towards FOSS platforms such as Python, R, and QGIS, which support open and reproducible workflows, is becoming mainstream.

Linked together in networks – whether the small ones developed by researchers to monitor air quality or water levels, or the large ones designed by firms to support mobile phone use or public transit – these systems capture *aspects* of the world in unprecedented detail: data – in their myriad forms as text, imagery, and operational records – are seen as key to unlocking a wealth of insight into the social and physical environment. However, in many cases researchers can only access these in an "accidental" manner (Arribas-Bel, 2014), implying that several of the channels, formats and quality checks scientists use with traditional data do not necessarily apply in this context. So to some (usually non-geography) proponents, the growth of 'big data' represents the 'end of theory' (Anderson, 2008), while to its detractors it represents a new kind of 'automated post-positivism' interested primarily in "selling you things that you don't actually need" (Wyly, 2014).

However accessible, the ultimate consequence of this reconfiguration of the data landscape is that the social sciences – and geography in particular – have gone from being data poor to being overwhelmed by a firehose of data sprayed towards us at high velocity, in high volumes, in a wide range of fast-changing formats, all while often being of dubious provenance (Kitchin, 2013). This trend is what has led some prominent geographical scholars to write of a 'data revolution' (Kitchin, 2014); other influential thinkers to go further and argue for the re-thinking of the methods and practices that researchers and analysts use to make sense of data, proposing a 'computational social science' (Lazer et al., 2009) or 'data science' (Donoho, 2017).

An emergent Data Science

> *I think statisticians are part of it, but it's just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively. But I do think those skills – of being able to access, understand, and communicate the insights you get from data analysis – are going to be extremely important.*
>
> *Hal Varian (2009)*

Burton (Burton, 1963, p.152) suggested that geography has long been a 'following discipline' whose "main currents of thought have had their origins in other fields." So, looking to our discipline's future, what currents are now taking hold? Who are we our should we be now following? Here we think it's worth turning to the emergent field of 'data science' and its use of algorithmic approaches to extract 'signal from noise'. Although the field is loosely defined (see Loukides, 2011; or Schutt and O'Neil, 2013 for illustrative attempts), competing disciplines, from statistics (e.g. Wu, 1997) to computer science (e.g. Naur, 1974), have sought to stake ownership of a terrain already occupied by the corporate behemoths of the early 21st Century.

In fact, in a paper derived from a commemorative speech in 2015 at Princeton, David Donoho (Donoho, 2017) traces the origins of contemporary data science back more than fifty years to John Tukey's *The Future of Data Analysis* (Tukey, 1962). Donoho's understanding, which seems to be one of the few formal attempts at synthesising what Data Science *is* without falling into marketing propositions or mere hype, points to a broadening of the traditional remit of statistics and, in particular, to the incorporation of six key components not traditionally taught as part of a 'statistics degree': data gathering, preparation, and exploration; data representation and transformation; computing with data; data visualization and presentation; data modeling; and a reflexive 'science of data science'.

Data science provides a framework to not only better understand, but also to effectively leverage 'data' (broadly defined), and this has created numerous, tangible advances in our capabilities to "do more with data". The reach of this emerging discipline spans hardware and software infrastructure, mathematical and statistical models, as well as methods and workflows. Specific examples include: using high-powered graphics cards to perform computation and massively parallel unstructured databases, open machine learning frameworks such as Google's TensorFlow (Abadi et al., 2016) and modern deep neural networks (LeCun et al., 2015), as well as the so-called tidy-verse of data manipulation and visualisation (Wickham, 2014; Grolemund & Wickham, 2017). So although it's possible to see Data Science as a largely improvisational and *ad-hoc* response by Silicon Valley startups to the need to deal with server logs that exceeded the available hardware and their lines of credit with specialised hardware vendors, the scope of application for Data Science is significantly greater.

Directly or indirectly, many of data science's applications are inherently spatial and geographic in nature, although the degree of engagement by what could be considered 'mainstream' data scientists with computational thinking originating in Geography has been fairly minimal. But within our discipline there is a widespread appreciation – built on the advances and struggles outlined above – that the majority of the behavioural data generated by our 'networked society' is spatially embedded and that geographical traditions may have much to offer 'big data' research. Everyone from Google and Airbnb to mobile phone carriers are in the geo-data business, and O'Sullivan and Manson (O'Sullivan and Manson, 2015) have, tongue planted firmly in cheek, suggested that this is one reason why physicists (amongst others!) are now the ones with geography envy.

Conclusion: Towards a Geographic Data Science

> *In the past decade geography has undergone a radical transformation of spirit and purpose, best described as the 'quantitative revolution'… Although the future changes will far outrun the initial expectations of the revolutionaries, the revolution itself is now over. It has come largely as the result of the impact of work by non-geographers upon geography…*
>
> *Ian Burton (1963)*

This paper has reviewed the relationship between computation and Geography since the invention of the modern computer in the 1950s; this can be briefly summarised as a journey that begins with the computer as an accelerator for analysing manually collected and tabulated data, to its emergence as critical not only to the analysis of data but also for its generation as well as a tool to create new knowledge through computational ways of *thinking*. We have also briefly reviewed the emergence of Data Science, largely at the interface between computer science and statistics, but incorporating ideas and practices that don't fit neatly into either discipline. And we have noted these have taken place largely at the periphery of Geography even if they are not completely foreign to the discipline. In this section, we would like to conclude by suggesting that there is a lot to gain from bringing these two strands closer together in what could become a Geographic Data Science (GDS).

What exactly a GDS should contain and how it should, pragmatically, be constructed is a task for a community, and not just two individuals. Our intention here is to start the discussion and stoke a debate. However, we would like to note that there is an urgency to building such a community since, in our view, the need for *something* like GDS is such that, one way or another and with or without geographers, it will be created. A principled refusal to engage with data science on epistemological or methodological grounds would leave parts of *our* disciplinary terrain and its (permeable) frontiers with other quantitative disciplines occupied by those with no appreciation of the history, the techniques, and the rationales underpinning spatially-aware quantitative analyses (see critical discussion in Brunsdon, 2014).

In parallel, there has been a 'hollowing out' of the skills required to make maps (Singleton, 2014); what used to be the preserve of those who had undertaken several years of 'GIS training' – voluntarily or otherwise – and had access to expensive proprietary software can now be done in a web browser. From data collection to data visualisation in the form of interactive, dynamic maps and the understandings that they enable, the entire pipeline can now be built without a geographer ever being involved. Geography has a long tradition of critical engagement with both data *and* their analysis and visualisation via GIS, and this awareness has a great deal to contribute to improving the kinds of insights that flow from sophisticated 'machine learning' and 'big data' approaches to spatial analysis.

Fortunately, it's not just about the risks as there is also much to be gained: Geography has always been a bridging discipline, connecting the natural and social sciences with the humanities; at the very least a GDS 'tradition' would provide a common ground between the computational geography and GIS traditions, and the emergent field of data science. Such a space for co-production would benefit both parties: computational geographers could collaborate with those who are pushing at the boundaries of what is possible to do with data and computers; data scientists could more easily benefit from a body of theory, practice, and expertise developed over decades that reflects on how space and location affects process and outcome, and that consequently seeks to explicitly account for such effects. Golledge has suggested that 'Geographers think differently' (Golledge, 2002 p.3) and the true value of a GDS may therefore resides as much in what we can bring to the table as what we can take away from it: it is not only about learning from each other, but also about developing spatially-integrated methods, tools, and techniques.

Ultimately, the challenges tackled by GDS may be entirely new and driven by access to novel forms of geo-data, or they may be more traditional questions which can now be tackled in entirely new ways. Crucially, GDS would use modern computational perspectives and expertise to fully exploit these data, while incorporating space – and the ethical and conceptual training of a social scientist – as an integral element of their production and interpretation. The understandings at which we might arrive are ones that neither group by themselves could arrive at on their own: a spatially-aware data science should be sensitive both to the substantive and insightful critiques of quantitative analyses mounted by cultural geographers, and to the ways in which 'data-generating processes' (Lu and Henning, 2012) are spatially determined. To put it another way: geographical data scientists understand *both* that latitude and longitude, Northing and Easting, and *x* and *y*, are just two dimensions amongst many in a big data set *and* that these axes retain a special 'power' over human experience.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Anderson, C., 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, *16*(7), pp.16-07.

Armstrong, M.P., 2000. Geography and computational science. *Annals of the Association of American Geographers*, *90*(1), pp.146-156.

Arribas-Bel, D., 2014. Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, pp.45-53.

Barnes, T.J., 2004. Placing ideas: genius loci, heterotopia and geography's quantitative revolution. *Progress in Human Geography*, *28*(5), pp.565-595.

Barnes, T.J., 2013. Big data, little history. *Dialogues in Human Geography*, *3*(3), pp.297-302.

Barnes, T.J., 2014. What's old is new, and new is old: History and geography's quantitative revolutions. *Dialogues in Human Geography*, *4*(1), pp.50-53.

Brunsdon, C., 2014. Spatial science–Looking outward. *Dialogues in Human Geography*, *4*(1), pp.45-49.

Brunsdon, C., 2016. Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography*, *40*(5), pp.687-696.

Burton, I., 1963. The quantitative revolution and theoretical geography. *The Canadian Geographer/Le Géographe canadien*, *7*(4), pp.151-162.

Cliff, A., 1977. Quantitative methods: Time series methods for modelling and forecasting. *Progress in Human Geography*, *1*(3), pp.492-502.

Cliff, A., 1979. Quantitative methods: spatial diffusion. *Progress in Geography*, *3*(1), pp.143-152.

Cliff, A.D., 1980. Quantitative methods: Spatial modelling. *Progress in Geography*, *4*(4), pp.568-576.

Cliff, A.D. and Ord, J.K., 1981. *Spatial processes: models & applications*. Taylor & Francis.

Couclelis, H., 1998, November. Geocomputation in context. In *Geocomputation: A primer* (pp. 17-29). John Wiley Chichester.

Donoho, D., 2015, September. 50 years of Data Science. In *Based on a Presentation at the Tukey Centennial Workshop*. Princeton: NJ.

Elwood, S., 2009. Geographic Information Science: new geovisualization technologies—emerging questions and linkages with GIScience research. *Progress in Human Geography*, *33*(2), pp.256-263.

Elwood, S., 2011. Geographic information science: Visualization, visual methods, and the geoweb. *Progress in Human Geography*, *35*(3), pp.401-408.

Fotheringham, A.S., 1997. Trends in quantitative methods I: stressing the local. *Progress in Human Geography*, *21*(1), pp.88-96.

Fotheringham, A.S., 1998. Trends in quantitative methods II: stressing the computational. *Progress in Human Geography*, *22*(2), pp.283-292.

Fotheringham, A.S., 1999. Trends in quantitative methods III: stressing the visual. *Progress in Human Geography*, *23*(4), pp.597-606.

Goddard, J.B., 1970. Functional regions within the city centre: A study by factor analysis of taxi flows in central London. *Transactions of the Institute of British Geographers*, pp.161-182.

Golledge, R.G., 2002. The nature of geographic knowledge. *Annals of the Association of American geographers*, *92*(1), pp.1-14.

Goodchild, M.F., 1991. Geographic information systems. *Progress in Human geography*, *15*(2), pp.194-200.

Goodchild, M.F., 2010. Twenty years of progress: GIScience in 2010. *Journal of spatial information science*, *2010*(1), pp.3-20.

Goodchild, M.F. and Haining, R.P., 2004. GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science*, *83*(1), pp.363-385.

Haining, R., 2014. Spatial Data and Statistical Methods: A Chronological Overview. In *Handbook of Regional Science*(pp. 1277-1294). Springer Berlin Heidelberg.

Harris, R., O'Sullivan, D., Gahegan, M., Charlton, M., Comber, L., Longley, P., Brunsdon, C., Malleson, N., Heppenstall, A., Singleton, A. and Arribas-Bel, D., 2017. More bark than bytes? Reflections on 21+ years of geocomputation. *Environment and Planning B: Urban Analytics and City Science*, *44*(4), pp.598-617.

Harford, T., 2014. Big data: are we making a big mistake? FT Magazine. *2014-03-28]. http://www. ft. com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0. html.*

Harvey, D., 1972. Revolutionary and counter revolutionary theory in geography and the problem of ghetto formation. *Antipode*, *4*(2), pp.1-13.

Kitchin, R., 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, *3*(3), pp.262-267.

Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. “Computational Social Science.” *Science* 323 (5915). American Association for the Advancement of Science: 721–23. doi:10.1126/science.1167742.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), p.436.

Ley, D. and Samuels, M.S. eds., 1978. *Humanistic geography*(p. 337). London: Croom Helm.

Longley, P.A., 2003. Geographical Information Systems: developments in socio-economic data infrastructures. *Progress in Human Geography*, *27*(1), pp.114-121.

Longley, P.A., 2004. Geographical Information Systems: on modelling and representation. *Progress in Human Geography*, *28*(1), pp.108-116.

Loukides, M., 2011. *What is data science?*. " O'Reilly Media, Inc.".

Lu, Y. and Henning, K.S., 2013. Are statisticians cold-blooded bosses? a new perspective on the ‘old’concept of statistical population. *Teaching Statistics*, *35*(1), pp.66-71.

Naur, P., 1974. Concise survey of computer methods.

Openshaw, S. 1994. “Commentaries.” *Environment and Planning A: Economy and Space* 26 (4): 499–508. doi:10.1068/a260499

Openshaw, S. and Abrahart, R.J., 2000. GeoComputation.

O'Sullivan, D., 2005. Geographical information science: time changes everything. *Progress in Human Geography*, *29*(6), pp.749-756.

O'Sullivan, D., 2008. Geographical information science: agent-based models. *Progress in Human Geography*, *32*(4), pp.541-550.

O’Sullivan, D. and Manson, S.M., 2015. Do physicists have geography envy? And what can geographers learn from it?. *Annals of the Association of American Geographers*, *105*(4), pp.704-722.

Phithakkitnukoon, S., Wolf, M.I., Offenhuber, D., Lee, D., Biderman, A. and Ratti, C., 2013. Tracking trash. *IEEE Pervasive Computing*, *12*(2), pp.38-48.

Poon, J.P., 2003. Quantitative methods: Producing quantitative methods narratives. *Progress in Human Geography*, *27*(6), pp.753-762.

Singleton, A. (2014). Learning to code. *Geographical Magazine*, p. 77.

Schutt, R. and O'Neil, C., 2013. *Doing data science: Straight talk from the frontline*. " O'Reilly Media, Inc.".

Thomas, R.W., 1990. Quantitative methods: input-output analysis as macroeconomic geography. *Progress in Human Geography*, *14*(3), pp.404-419.

Torrens, P.M., 2010. Geography and computational social science. *GeoJournal*, *75*(2), pp.133-148.

Tukey, J.W., 1962. The future of data analysis. *The annals of mathematical statistics*, *33*(1), pp.1-67.

Unwin, D.J., 1995. Geographical information systems and the problem of'error and uncertainty'. *Progress in Human Geography*, *19*(4), pp.549-558.

Unwin, D.J., 1996. GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, *20*(4), pp.540-551.

Wickham, H., 2014. Tidy data. *Journal of Statistical Software*, *59*(10), pp.1-23.

Wickham, H. and Grolemund, G., 2016. *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".

Wrigley, N., 1982. Quantitative methods: developments in discrete choice modelling. *Progress in Geography*, *6*(4), pp.547-562.

Wrigley, N., 1986. Quantitative methods: the era of longitudinal data analysis. *Progress in Human Geography*, *10*(1), pp.84-102.

Wu, J., 1997. Statistics = Data Science? Available at https://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf

Wyly, E., 2014. Automated (post) positivism. *Urban Geography*, *35*(5), pp.669-690.